DETECTION OF EPILECTIC SEIZURES FROM INTRACRANEAL EEG DATA

Bachelor Thesis in Biomedical Engineering
June 2015

Author: Eva García Alonso
Tutor: Jose M. Leiva Murillo

Universidad Carlos III de Madrid
Escuela Politécnica Superior

Acknowledgements

Primero de todo, me gustaría agradecer a Jose Miguel Leiva su dedicación, disposición y atención desde el primer momento.

También me gustaría darle las gracias a mis compañeros y amigos Isabel, Alba y Ander por su paciencia, por haberme acompañado y apoyado a lo largo de estos 4 años. Sin ellos, esto no hubiera sido posible.

A Bea y Rubén por nunca dudar de mis posibilidades incluso cuando yo lo hacía.

Además, agradecer a mi segunda familia en Galicia por estar siempre presentes aún a tantos kilómetros.

Por ultimo y más importante, a mis padres y mis hermanas por enseñarme que el trabajo duro vence todo obstáculo que se cruce. Por ser mi ejemplo e inspiración en mi vida.

INDEX of CONTENTS

INDEX of EQUATIONS

INDEX of FIGURES

1. INTRODUCTION

Epilepsy is one the most common neurological diseases that affects 0.5-1% of the population of the world [1]. It is characterized by seizures that could or could not have some periodicity in its repetitions.

On the other hand, although there exist a high variety of antiepileptic drugs, approximately between 30% and 40% of the patients that suffer this disease could not be controlled only by these drugs [2]. In addition, most of the antiepileptic drugs have some side effects like hypertension, diabetes o heart disease as well as some cognitive side effects. [3]. Moreover, the quality of life of the patients is decreased due to the uncertainty on when a new seizure will occur or their continuous revisions in order to adapt the medication to new possible variety of symptoms. Although the main reason of the mortality associated with epilepsy are the consequences of unconsciousness, preventing and detecting the possible incoming seizures could provide the solution to the all these problems.

Some researchers support the idea that the ictogenesis, that is the process of seizure generation, is not random. It is related with EEG and its characteristic pattern that allow us to study the relation between it and the epilepsy attacks. In addition they also support that some changes occur in the brain before epileptic seizures occur.

This Bachelor Thesis presents a new and innovative proposal for epileptic patients that have not the expected results only with medication. The system consists on a device that provides an intracranial electroencephalogram that is a continuous, real-time measurement of the brain activity and brain waves. This simple device placed on the surface of the brain is connected with an advisory system. However, the part in which this Bachelor Thesis is based on is the one related with electroencephalogram data processing. In this way, it opens the

possibility of the detection and forecast of an incoming seizure and leaves enough time to allow doctors to act to a possible threat for the patient.

In addition, another problem that the antiepileptic drugs present is that they are not a specialized treatment for each of the patient, whose epilepsy could be different, and in consequence, the medication will not act exactly in the same manner in all the patients. This innovative system will provide an individual treatment for each patient, acting only when necessary for each of the individuals of the study.

Thus, the idea is to study the data of intracranial electroencephalogram (iEEG) that is implanted in 4 dogs and 8 patients in order to get those patterns and get some periodicity between them to detect the attacks. This iEEG uses algorithms to detect the seizures and deliver responsive stimulation to avoid them. In addition, it allows us to investigate the origin of the seizures. [4]

The first objective of this Bachelor Thesis is to create an algorithm that allow us to classify a set of data (test data) into ictal regions (epilepsy attacks) and interictal regions (non epilepsy attacks) on the basis on another set of data that has its own classification (training data). In order to do that, some concepts about logistic regression and cross-validation will be essential.

Once this problem is solved, the second aim is to separate the first fifteen seconds of the attack (early ictal segments) from the rest of the attack segments (late ictal segments). Detecting early ictal segments will provide the possibility of acting against an incoming seizure before it occurs or at the very beginning of the epilepsy attack. In addition, assessing future health care benefits as well as a study of those benefits will be achieved with this project.

The software used for these objectives will be Matlab_R2014a software that is a mathematical tool that allows to apply all these statistical principles previously appointed like logistic regression, characterization of signals or data analysis.

In the following chapters, the way to detect and control these seizures will be explained as well as previous knowledge that is necessary to understand how and why the device is implanted and in which mechanisms it is based on.

2. BACKGROUND

Epilepsy is a complex process and mechanism that occur spontaneously. In order to understand the mechanism of the process, it is necessary to have some prior learning about how the brain works (neurophysiology) as well as knowledge about data analysis and statistical or mathematical processes for classification.

2.1 Neurophysiology

Neurophysiology is the branch of the physiology dealing with the nervous system. At the same time, physiology is the branch of biology dealing with the functions and the activity of the organs and systems of the body, including all physical and chemical processes.

In this section, there are some important concepts that are going to be introduced and that have relevancy and importance with the concern matter previously exposed. These topics are the nervous system; the electrical activity in neurons; brain waves and electroencephalogram and finally epilepsy.

2.1.1 NERVOUS SYSTEM

The nervous system, with just 3% of the total weight of the body, is one of the smallest and yet most complex of the body systems. It is a complex collection of nerves and specialized cells known as neurons that transmit signals between different parts of the body [5]. Along with the endocrine system, it is responsible of maintaining controlled the conditions for homeostasis but also it is responsible of memories, movements and behaviors.

The nervous system has three main basic functions:

- Sensing changing with the sensory receptors. The responsible neurons of sensing are the afferent or sensory neurons.
- Interpreting and remembering those changes. The responsible neurons of that are the interneurons.
- Reacting to those changes with effectors (muscular contractions and glandular secretions). The responsible of the motor function are the efferent neurons.

The nervous systems is divided into central nervous system (CNS) and peripheral nervous system (PNS):

- The central nervous system (CNS) is composed of the brain and the spinal cord.
- The peripheral nervous system (PNS) consists on cranial and spinal nerves that contain both sensory and motor fibers, plus ganglia and plexures. At the same time, the PNS can be divided into: somatic, autonomic and enteric nervous systems. First, somatic NS is responsible of motor neuron of the skeletal muscle tissue and special sensory receptors to the CNS. Second, autonomic NS is responsible of motor neurons of cardiac muscle. Finally, enteric NS is responsible of motor neurons of the gastrointestinal tract. [6]

2.1.2 ELECTRICAL SIGNALS IN NEURONS

Neurons are electrically excitable due to voltage differences across the membrane. There are two main types of potentials or electrical signals: action potentials that travel at long distances and graded potentials that are local membranes changes.

For our purpose, the electrical signal in which we are interest is the action potential. The action potential is a sequence of events that provokes the resting potential to reverse the membrane potential (depolarization) and then restore it to the resting potential (repolarization). According to the *all-or-non principle* the

if a stimulus reaches threshold, the action potential is always the same, that is, a stronger stimulus will not cause a larger impulse. [6].

First of all, the resting potential is the static potential of the membrane if any change is produced (stimulus) that is -70 mV since the concentration of ions inside and outside is different. During the depolarization phase, the action potential rises to +30 mV and then is restored to -70mV during repolarization. After that, there exists a refractory period that is a period of time after an action potential begins during which an excitable cell cannot generate another action potential. [6]

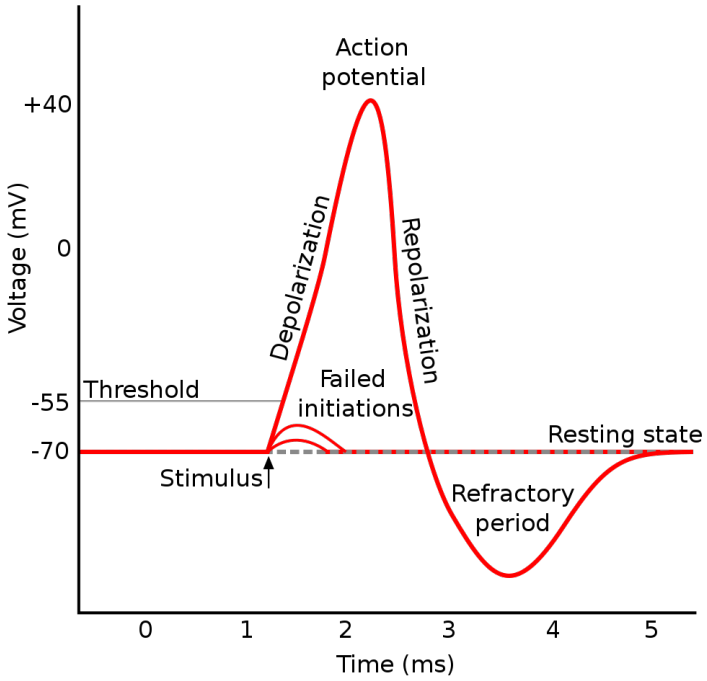

Figure 1: Action Potential [6]

The action potential travels along the membrane of the axon of the neuron and is called nerve impulse. This nerve impulse is transmitted from one neuron to another or from one neuron to an effector organ by a process called synapsis.

The propagation speed is not related with the stimulus strength. The way to differentiate a weak stimulus from a strong stimulus is by the frequency of impulses: the higher the impulse, the higher the frequency.

2.1.3 BRAIN WAVES AND ELECTROENCEPHALOGRAM (EEG)

The electroencephalogram (EEG) is a measurement of electrical activity of the brain or brain waves. It is a test that provides information about how the brain works over time. It is used for diagnosis of brain disorders as well as for introducing human-machine interface to improve patient's life.

The recording of the electrical activity demonstrates that there is a continuous electrical activity in the brain. Both the intensity of such electrical activity and its patterns allow experts to determine wherever a patient is asleep, awake or some brain diseases such as epilepsy or even psychoses. These patterns are called brain waves.

These waves are very difficult to relate with a specific pattern, sometimes it is possible and sometimes it is not. In healthy patients, most of the waves in the EEG are classified as alpha, theta, beta and delta waves and all are present in the electroencephalogram at the same time that makes even more difficult to classify them and to find repetitive pattern.

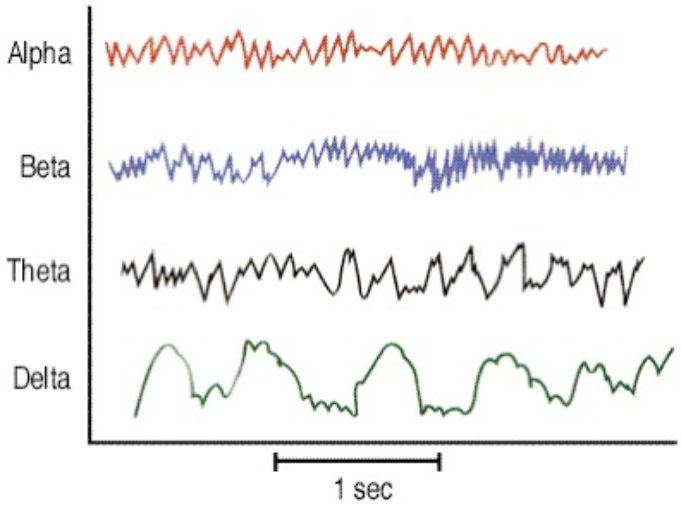

Figure 2: Brain waves [6]

- Alpha waves occur at frequencies between 8 and 13 cycles per second (Hz) and are characteristic of adults that are awake in resting state cerebration. Their voltages are approximately +50 μV.
- Beta waves occur at frequencies between 14 and 80 cycles per second. They are characteristic from the parietal and frontal specific activity.
- Theta waves occur at frequencies between 4 and 7 cycles per second. They are characteristic of parietal and temporal regions of children and in some specific cases of emotional stress in adults. It is an important characteristic of some brain disorders.
- Delta waves occur at frequencies lower than 3.5 Hz and with much higher voltage in comparison with the other waves. They are characteristic of the deep sleep in infancy and in important brain diseases. [7]

During different steps of the wakefulness and sleep, the activity of each wave is very characteristic as it is shown in the Figure 3.

In this Bachelor Thesis instead of using extracranial EEG, intracranial one was used. Intracranial EEG is the direct measurement of the brain waves from the brain surface which needs an invasive surgical process to place the electrodes to the scalp. Intracranial EEG is often called electrocorticography. By applying this kind of electroencephalogram, some movement's artifacts are avoided and allow to have a measurement that has more sensitivity in the exact area where the attack occurred.

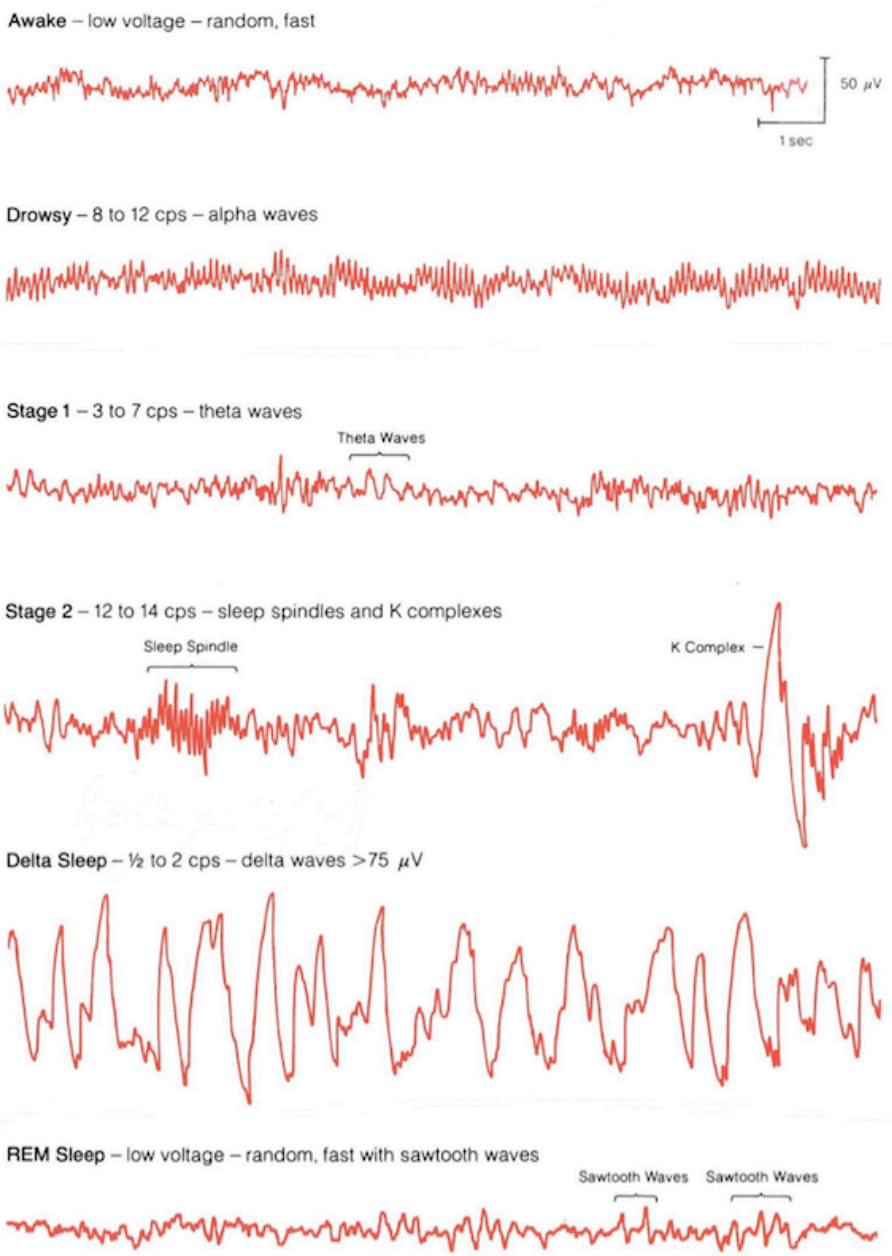

Figure 3: Brain waves during different stages of sleep and wakefulness [6]

2.1.4 EPILEPSY

Depending on the activity and in consequence the pattern of the brain waves, some brain diseases could be diagnosed. One of these disorders that can be detected is epilepsy. Epilepsy is characterized by uncontrolled excessive activity of either part or all the central nervous system. [7]. The epilepsy seizures occur when the level of excitability rises above a threshold. Epilepsy can be classify into 3 main classes:

- Grand Mal Epilepsy. Extreme discharges in all areas of the brain sometimes reaching the spinal cord causing tonic seizures of the whole body. This kind of epilepsy could even affect respiratory system. It lasts from a few seconds to 3 or 4 minutes.
- Petit Mal Epilepsy. Affect the thalamocortical brain activating system. It is characterized by 3 to 30 second of unconsciousness in which some muscle contractions occur. The pattern of the petit mal epilepsy could be seen in the Figure 4.
- Focal Epilepsy. Can occur along any region of the cerebral cortex or deeper structures or both the cerebrum and brain stem. This type of epilepsy usually occurs due to some localized lesion of functional abnormality.

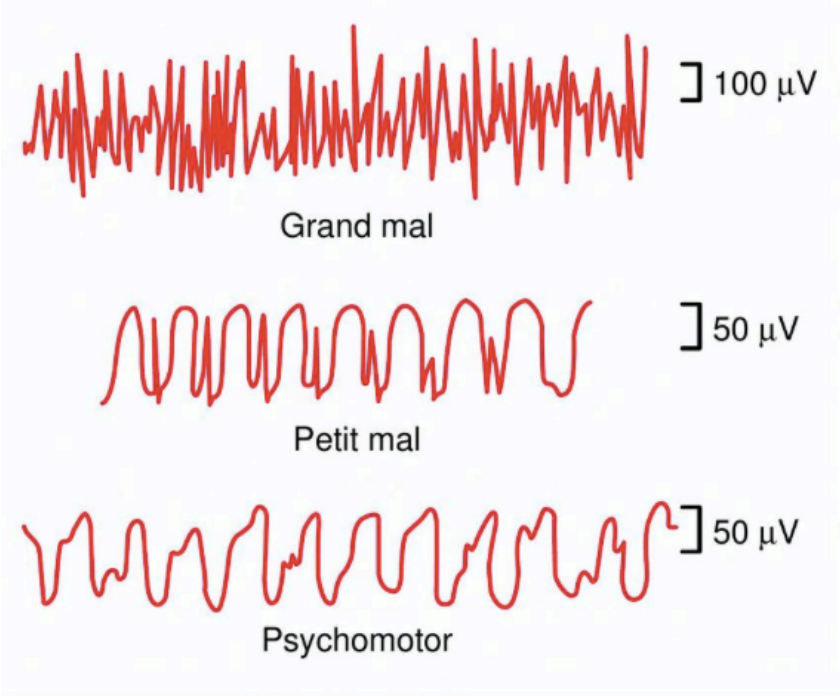

Figure 4: Grand Mal, Petit Mal and Focal epilepsy patterns [6]

About 1% of people worldwide suffer epilepsy disease and approximately 80% of cases occur in developing countries. [8]

In the developed world, epilepsy is more frequent in infants and the elderly while in the developing world it is more frequent in older children and young people. [9]

2.2 Signal and Data Processing

A broad definition of data processing is “the collection and manipulation of items of data to produce meaningful information.” [10]

Data processing allows to fulfill any of these applications:

- Validation
- Sorting
- Analysis
- Reporting
- Classification

In this Thesis I am going to focus in analysis of data processing in such a way that given some raw data I am going to convert it into useful information for the user. These raw data is commonly known as big data that is a set of data that is very complex and impossible to manage.

In addition, there are two kind of analysis in data processing: quantitative analysis and qualitative analysis. The first, the one that we are dealing with, is based in the collection and analysis of quantitative data through variables while qualitative data avoid quantification. The qualitative researchers make narrative registers of phenomena through participant observation and unstructured interviews.

In order to achieve this goal, the quantitative data analysis process has several steps:

- Data collection: the best choice is to select wide data from diverse sources in such a way that the possibility of finding and building the best models in as high as possible.
- Data cleaning: this is the most important and critical step. It is vital a good selection of the data by deleting these that has errors or nonsense information.
- Data modeling: it is essential to have a good background in statistics and machine learning to achieve the best model.
- Optimize and repeat

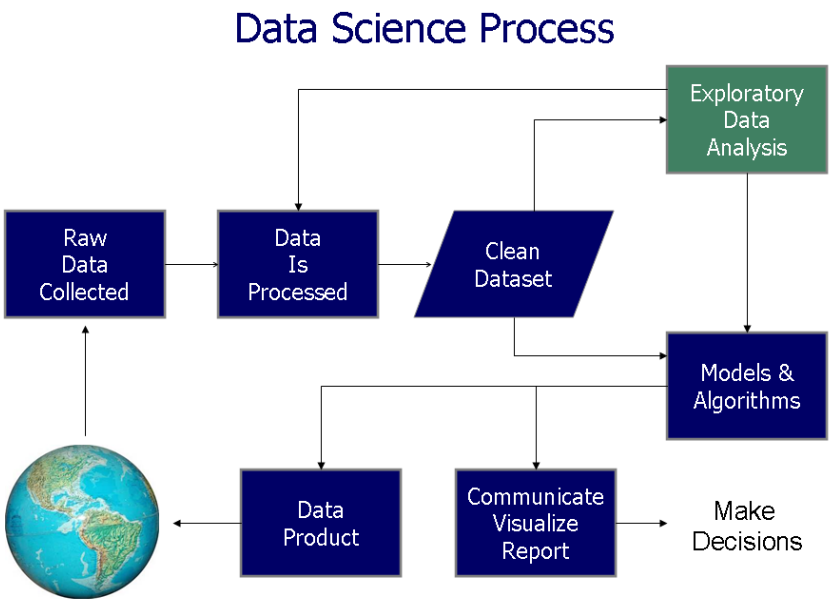

Figure 5: Data Process [10]

To process and classify the electroencephalogram signals, the characterization of them is an important process to find some relations and similitudes between some patterns and another.

2.2.1 STATISTICAL CHARACTERIZATION OF SIGNALS

Signals are physical manifestations of natural or artificial processes of different nature. They are related to one or more independent variables and they content information about behavior of some physical phenomena.

In order to characterize signals there are some important parameters that give the necessary information to allow the data analysis.

- Power and energy. Power gives information about the consumption that is done to transmit a signal. In this way, a signal with higher power could be sent to farer places than a signal with lower power. So, the power gives information about the energy of a signal per unit of time and is measured in Watts. On the other hand, energy is timeless. Thus, given a signal x(t):

$$E = \int_{-\infty}^{\infty} |\, x(t)^2 |\, \delta t$$

Equation 1: Energy

$$P = E(|\, x(t)^2 \,|)$$

Equation 2: Power

  The energy of a periodic signal is infinite while the power of a aperiodic signal is infinite. Thus, we are talking about energy signals when the energy has a significant value and power is infinite while power signals are those that have a significant value of the power and an infinite value of the energy.

- Covariance. It is a value that indicates the intensity of the total variation between two random variables. It is an important parameter to determine the dependency between two variables and it is very useful for the determination of other parameters.

$$\sigma(x,y) = E[(x - E[x])(y - E[y])] = E[xy] - E[x]E[y]$$

Equation 3: Correlation between two variables x(t) and y(t)

- Standard deviation. It shows the dispersion or variety of a distribution of quantitative or qualitative values.

$$\sigma^2 = \int (x-\mu)^2 f(x)\ \delta x \quad \text{where} \quad \mu = \int x f(x)\ \delta x$$

Equation 4: Standard deviation of a signal x(t)

- Correlation coefficients. They indicate the force and direction of a linear relation between two statistical variables. Given two quantitative variables, they are correlated when the values of one of them vary with respect to the values of the other, that is, if we have two variables there exist correlation between them if when increasing the values of one of them, the values of the other also increase or vice versa.

  There exist different coefficients that measure the correlation between two variables. The one that is most known is Pearson's coefficient that divides the covariance of both variables by the product of their standard deviations.

$$\rho(x,y) = \frac{\sigma(x,y)}{\sigma_x \sigma_y} = \frac{E[(x-\mu_x)(y-\mu_y)}{\sigma_x \sigma_y}$$

Equation 5: Correlation

- Eigenvalues. They are closely related to the value of the correlation coefficients. In some manner, N eigenvalues summarize $\frac{N^2}{2}$values of the correlation coefficients, so it is a way to represent the coefficients in a concise manner. Moreover, the spectrum of the eigenvalues will also give information about correlation (if the spectrum provides high energy in low number of eigenvalues, the correlation is much higher and vice versa).

- Power Spectral Density. In contrast with the power, the power spectral density shows the dispersion of the power along the frequency.

$$P_x = \int_{-\infty}^{\infty} S_{xx}(f)\, \delta f$$

Equation 6: Relation between power and PSD

where S_{xx} is the power spectral density.

So,

$$S_{xx} = \lim_{T \to \infty} \frac{|X(f)|^2}{T}$$

Equation 7: Power Spectral Density

2.3 Machine Learning

Machine learning is a branch of the artificial intelligence that develops some techniques to allow machines to learn. In general terms, these systems are capable of analyzing the environment and learn from it. In a technical way, the method tries to create programs that are capable of generalize some behavior by analyzing some info. So many times, the field of machine learn is very related with the statistics fields since both techniques are based in data analysis. However, the automatic learning or machine learning could be seen as a try to automatize some parts of the scientific method through mathematical methods like statistics. Some systems that form parts of machine learning are neuronal networks, support vector machines and logistics regression.

My work was centered in the classification process. Taking some samples with a determined label, a classifier was trained to achieve the division of some test samples into two different classes.

2.3.1 SUPPORT VECTOR MACHINES

The SVM consists on a group of learning algorithms implemented to solve problems related with classification and regression. The system follows some ordered steps. Firstly, we need some data samples that are classified in one category or in another. Then, with these data the support vector machine is trained to build a model that allows them to classify some test data in one of both classes. What it exactly does is to create a hyperplane or a group of them in high dimensions, even infinite dimension, which could be used in classification and regression problems. This kind of algorithms tries to select the hyperplane that maximizes the distance between this hyperplane and the points that are closer to it. This process is explained with a simple 2-dimensional example in Figure 6.

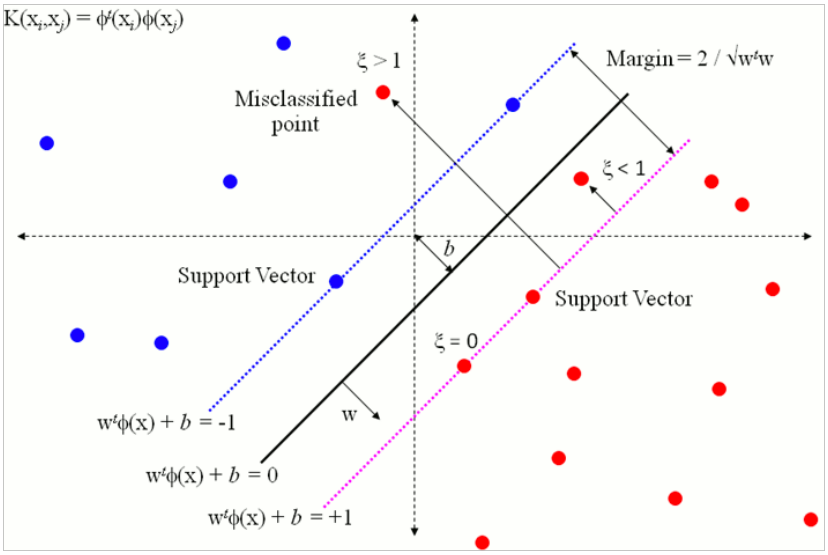

Figure 6: Suport Machine Vector in 2D [11]

2.3.2 LOGISTIC REGRESSION

Logistic regression is one form of machine learning that is used in order to predict the result of a categorical value (the one that could adopt a limited number of categories) as a function of the independent variables. It is the one that we are going to use in order to train the classifier.

This kind of regression is used in order to obtain an output that could only be two values: 0 or 1 (or in our case, -1 and 1). This kind of data is very useful in biomedicine to predict different diseases or answers to some drug. In this way, a positive prediction is represented with "1" and a negative prediction is "-1". The main objective is to research about the relation between the probability related to the output variable and the independent input variables.

In these conditions, the probability would be calculated through a logistic function:

$$P(y=1|x) = \frac{e^{(b+\boldsymbol{w}^T\boldsymbol{x})}}{e^{(b+\boldsymbol{w}^T\boldsymbol{x})}+1} = \frac{1}{e^{-(b+\boldsymbol{w}^T\boldsymbol{x})}+1} = \frac{1}{e^{-y}+1}$$

Equation 8: Estimated Probability for y=1

$$P(y=-1|x) = 1 - P(y=1|x) = \frac{1}{e^{y}+1}$$

Equation 9: Estimated Probability for y=-1

$$y(x) = ln\frac{P(x)}{1-P(x)} = \boldsymbol{w}^T\boldsymbol{x} + b$$

Equation 10: Logistic Regression

Where x is the independent data, P the estimated probability, y is 0 or 1 depending on each case and g is the logistic regression.

Although the general formula of the probability is the one represented in Equation 8, this comes from maximizing the log-likelihood in order to calculate the "w" parameter. Thus, in general:

$$\begin{aligned} \mathrm{L(w)} = \max \prod_i P(y_i|x_i) = \max\ [\log \prod_i P(y_i|x_i)] = \max\ [\sum_{i=1} \log P(y_i|x_i)] \\ = \max\ [-\sum_{i=1} \log(1+e^{-\boldsymbol{w}^T\boldsymbol{x}_i})]\ = \min\ [\sum_{i=1} \log(1+e^{-\boldsymbol{w}^T\boldsymbol{x}_i})] \end{aligned}$$

Equation 11. Maximization of the log-likelihood

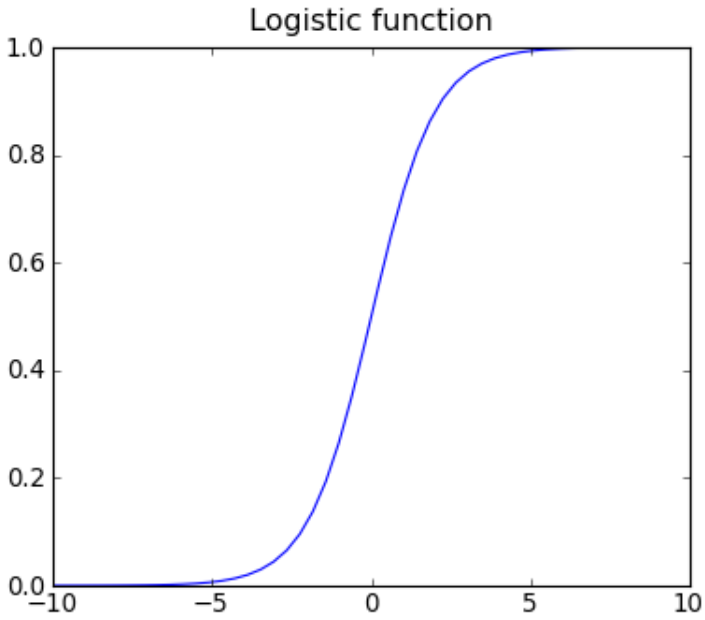

Figure 7: Logistic Regression Function

Although this method is simple to use, it has one problem: the matrix "X", that is the matrix that accumulated all the vectors **x** of each of the segments on one individual, could be singular, that is, a matrix that has not inverse and that its determinant is 0. This is a problem because in order to calculate "**w**" it is needed the inverse of "**x**".

This problem is called Weighted Least Square Problem and the solution implemented in my algorithm is the Newton's method that is called Iterative Reweighted Least Square [12]:

$$L(w) = \min\ [\sum_{i=1} \log(1 + e^{-wx_i})]\ \ + \lambda\|w\|^2$$

Equation 12: Maximum log-likelihood with parameter λ

Graphically, L(**w**) is represented in Figure 8. Since it is a convex representation, it means that the local and total minimum will be the same. With the parameter λ what we are trying to do is to optimize the logistic regression by finding the value of λ that makes L(w) to be as closer to the minimum as possible. Moreover, if we put a value of λ that is very low or very high, the

algorithm will not be optimized. With a very small value, λ is not even smoothing the algorithm and with very high values, the weight of λ will be as high that it will be the only parameter affecting the algorithm.

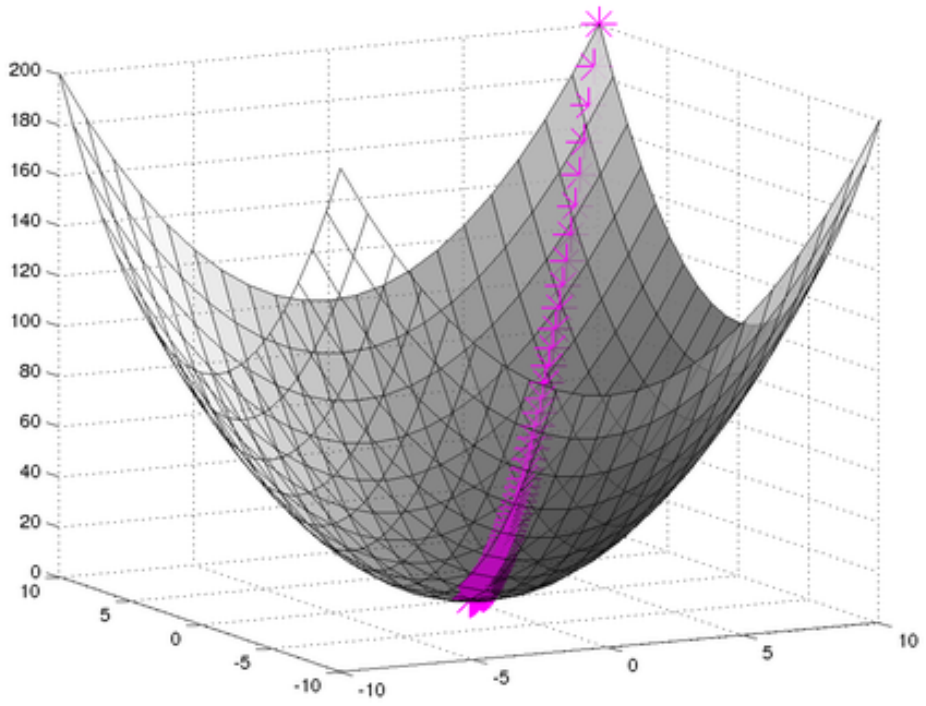

Figure 8: Convex function of L(w)

2.4 Cross-validation and ROC Curve

Cross-validation is a technique used to evaluate the results of a statistical analysis to guarantee that these results are independent of the partition between training and test data. During k-fold cross-validation, training data is divided into k subgroups. One of these subsets is used as test data and the rest of the subgroups as training data. It is repeated k times changing each time the subgroup that is responsible of testing.

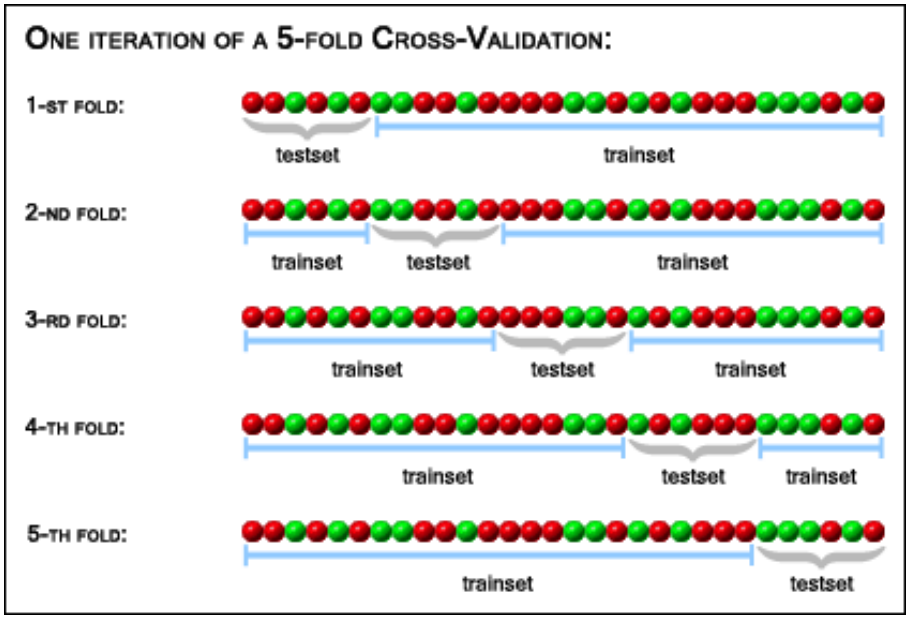

Figure 9: Cross-validation [13]

Although it is a very precise and efficient method, its main drawback is that it is a very slow process for the programming and computational point of view.

However, in order to determine how good the algorithm proposed is, we need to compare the data from the cross-validation with the real label of the segments of the training data. This determination is done with a Receiving Operating Characteristics curve (ROC curve) that is a representation of the sensitivity versus (1-specificity). Another option of interpretation of the graph is the ratio true positive versus false positives.

Consider a general prediction problem like the one of this bachelor thesis where the results are classified into positives (p) or negatives (n). There are four possible results: if the result is positive and the true label is positive, then it is known as true positive; if the result is positive and the true label is negative, it is known as false positive; if the result is negative and the true label is positive, it is known as false negative; finally, if the result is negative and the true label is negative too, it is known as true negative. All these four values are collected in the confusion matrix shown in Figure 10.

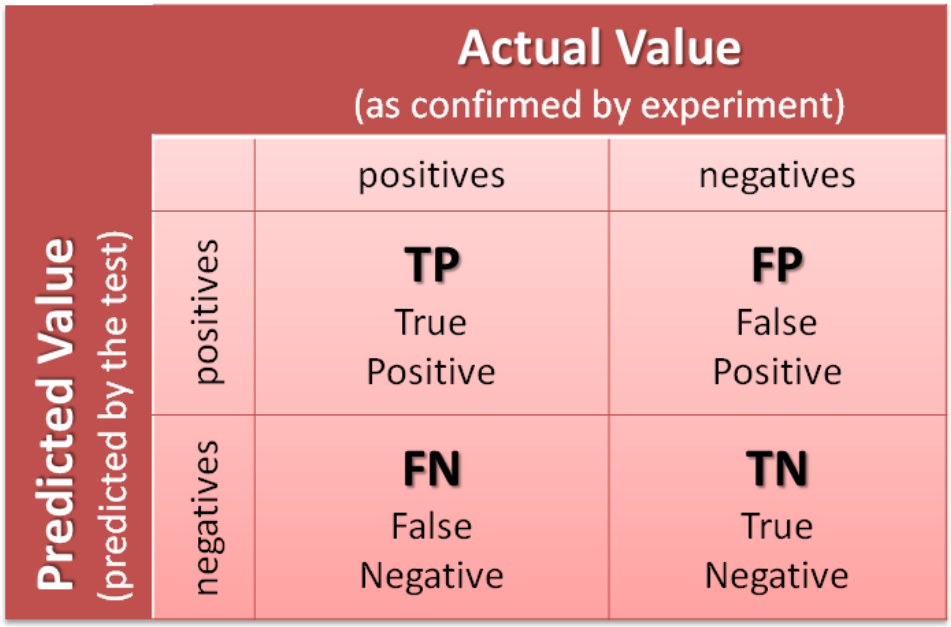

Figure 10: Confusion Matrix

Anyway, to draw the ROC curve the only parameters needed are the true positive and false negatives. On one hand, true positive measures how our classifier is able to detect the positive samples correctly. On the other hand, false positive measures how many samples are classified as positive incorrectly. So, the ROC curve represents in the x-axis the false positive and in the y-axis the true positives. So, since true positive detection is the same as sensitivity and false positive is the same as (1-specificity), the ROC curve is also known as sensitivity vs (1-specificity) as I have said previously.

Based on that, the best possible prediction will be situated in a point of the left up side (coordinate (0,1)) representing 100% of sensitivity and 100% of specificity.

Red discontinuous line shown in Figure 11 represents a random classification. A typical example of the random guess would be to decide the results from the result of flipping a coin. So, all the results must be above this line if they are, at least, a bit better than flipping that coin.

In order to known the exact percentage of the coincidence with the real labels of the training data, it is as simple as calculate the area under the ROC curve.

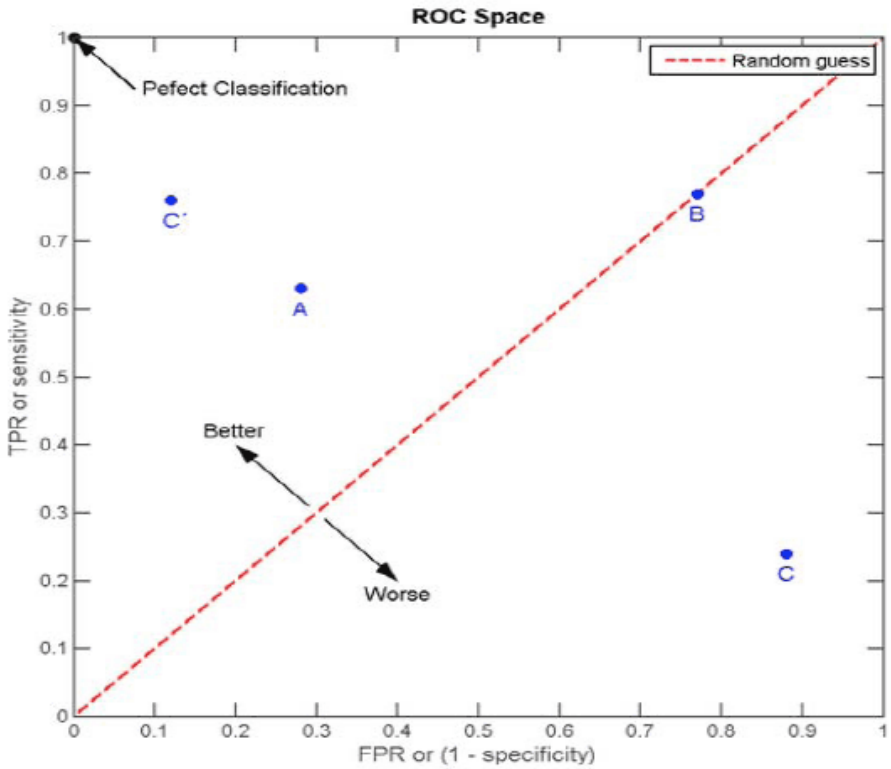

Figure 11: ROC curve [14]

3. PREPROCESSING

In this chapter, although the main objectives were not these, a brief explanation of the device and the preprocessing that was carried out is going to be done to completely understand the process for a future use.

3.1 Device

An implantable iEEG acquisition system was implanted to measure a long-term continuous EEG of 4 dogs and 8 patients. The device has 3 main components:

- An Implantable Lead Assembly (ILA). It is composed on silicon and platinum-iridium contact separated by 20 mm. This part of the device takes the signals from the iEEG. It is placed in the brain, exactly in the region that is believed to be the responsible of the epileptic attacks (this is determined with previous imaging and EEG studies). It is similar to those leads used in epilepsy monitoring units but that are more robust and more tolerant to stresses.
- An Implantable Telemetry Unit (ITU). This part filter, amplify and digitize the signal. It is placed in the subclavicular region.
- Personal Advisory Device (PAD). It is responsible of the wireless transmission. [4]

Figure 12 represents the different parts of the device implanted in a dog and the main uses of each of them.

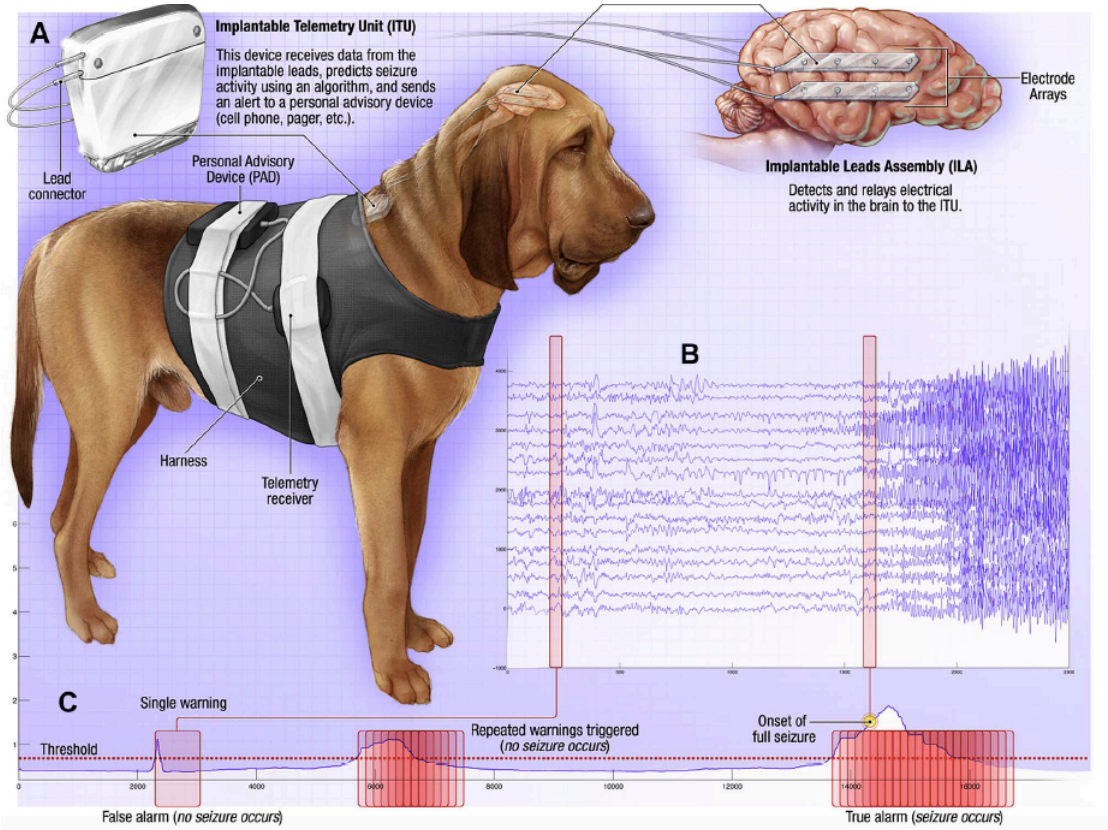

Figure 12: Seizure Device [4]

3.2 Database and Filtering

The data and idea for this bachelor thesis was extracted from UPenn and Mayo Clinic's Seizure Detection Challenge proposed by the University of Pennsylvania and the Mayo Clinic [15]. The given data of the problem consist on a set of training samples and test samples. The training samples are divided into two classes: ictal, which is the one that represents an epileptic seizure, and interictal, that is the one that represents the samples that are not in an epileptic seizure. In this way, this work tries to train the classifier to create hyperplanes that separate the test data into ictal or interictal in the most efficient way. On the other hand, the project has a second part in which the aim is to separate the test samples into early or late epileptic seizure. Early samples are those that belong to the fifteen first seconds of the ictal section and late samples are the ones that belong to the rest of the ictal section. So, in this part, the classifier was trained with the training data to allow us to separate the test data into early and late,

that is, interictal and late samples are classified as one category and early as another category.

It is in this part where logistic regression plays an important role. Logistic regression will allow us to do this classification in an efficient way.

Each file is composed of raw data for 1 second and depending if it is a dog or a patient it has different number of channels. For each dog there are 16 channels per iEEG that is continuously recording and wireless transmitted to the PAD and then stored in a flash memory card. On the other hand, the human patients have not got the same number of channels per iEEG in all of the patients. The number of channels of the iEEG goes from 16 to 72 depending on each patient.

Artifacts, especially those related with the ocular movement, seem to be the main drawback with which researchers and doctors have to deal everyday. Although the artifacts in intracranial EEG are not as important as in extracranial EEG, the data that intracranial EEG provides is also subjected to some artifacts that would affect the processing of the signals during training and testing. Although in our the data is already filtered and this preprocessing is not necessary, one of the most important artifacts is the power line interference at 50 and 100 Hz cycles that is removed by applying a band-pass filter that eliminates one region between 47-53 Hz and another between 97-103 Hz. On the other hand, the bipolar recording method is used to eliminate the movement artifacts. [16].

The given data of the problem consist on a set of training samples and test samples. The training samples are divided into two classes: ictal, that is the one that represents an epileptic seizure, and interictal, that is the one that represents the samples that are not in an epileptic seizure. In this way, this work tries to train a classifier to create hyperplanes that separate the test data into ictal or interictal in the most efficient way. On the other hand, the project has a second part in which the aim is to separate the test samples into early or late epileptic

seizure. Early samples are those that belong to the fifteen first seconds of the ictal section and late samples are the ones that belong to the rest of the ictal section. So, in this part, the classifier was trained with the training data to allow us to separate the test data into early and late, that is, interictal and late samples are classified as one category and early as another category.

It is in this part where logistic regression plays an important role. Logistic regression will allow us to do this classification in an efficient way.

4. PROPOSED SOLUTION

Detection of epileptic seizures is a process that has several steps and a huge investigation about the independent variables that during the logistic regression are essential for the aim. During the process we are following, what we do is to compare the different spectrum and variables between the channels (different positions of the electrodes in the brain) in order to find some change in both at the same time that allow us to classify them into ictal or interictal and early or late.

4.1 Interictal vs Ictal

Classifiers based in logistic regression are the ones used for this purpose. They are applied in Matlab_R2014a software. The aim is to train the classifiers to differentiate the label of interictal or pre-ictal (y=-1) and the ictal ones (y=1). In order to achieve the goal, the process followed for dogs and patients is almost the same but I am going to explain them separately to remark the differences between them.

4.1.1 DOGS

The data is structured in the following way: for each individual we have hundreds of windows of one second. Each of them contains 4 files: data of the iEEG, latency in the case of the ictal files that provides us the information regarding the exact position of each segment during the epileptic attack to allow us to differentiate into early and late, a data called channel that provides the number of channel that each segment and individual has and the sampling frequency. In dogs, the number of channels is constant and is always 16. The sampling frequency is also the same, that is, 400 Hz. Moreover, neither the number of files nor the number of epilepsy attacks for each individual is not constant.

During logistic regression we need some independent variables that are needed to train the classifier. The variables used are correlation between channels, power spectral density, eigenvalues of the correlation between channels and power of the signal.

In order to calculate the power spectral density we need to take into account that if we take the 400 samples for the calculation of it, the resulting vector will have a very high dimension difficult to manage and that will slow down the piece of code. However, the most important reason is that for our purpose, by calculating the PSD of 50 samples and calculating the mean will provide a more reliable value. With the purpose to decrease the dimension of the PSD vector, I have calculated the PSD for 50 samples eight times and I have calculated the mean. In this way, the PSD matrix is going to be reduced into dimension 50 instead of dimension 400. This method is called the Barlett's method. [17]

The selection of those variables is closely related with the similarity of the iEEG between channels. First of all, correlation is an important factor because during interictal process (no seizure occurs) the activity of channels is very similar between them (so high correlation) but when a seizure occurs, this correlation changes and decreases because the attack is more active in some zones than in others [16]. Secondly, power is altered during a seizure, that is, the power is much higher when an epileptic attack is occurring. The same seems to happen with power spectral density, during a seizure the spectrum has much higher values of high frequencies than during interictal events.

Finally, eigenvalues are closely related with the value of the correlation. During interictal events, the correlation is much higher and it is represented like in Figure 13. In consequence the matrix is the following:

$M = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$ therefore the eigenvalues are $\lambda = 2 \; and \; \lambda = 0$.

So, that means that, during the interictal segments the energy of the signal is accumulated in some few eigenvalues.

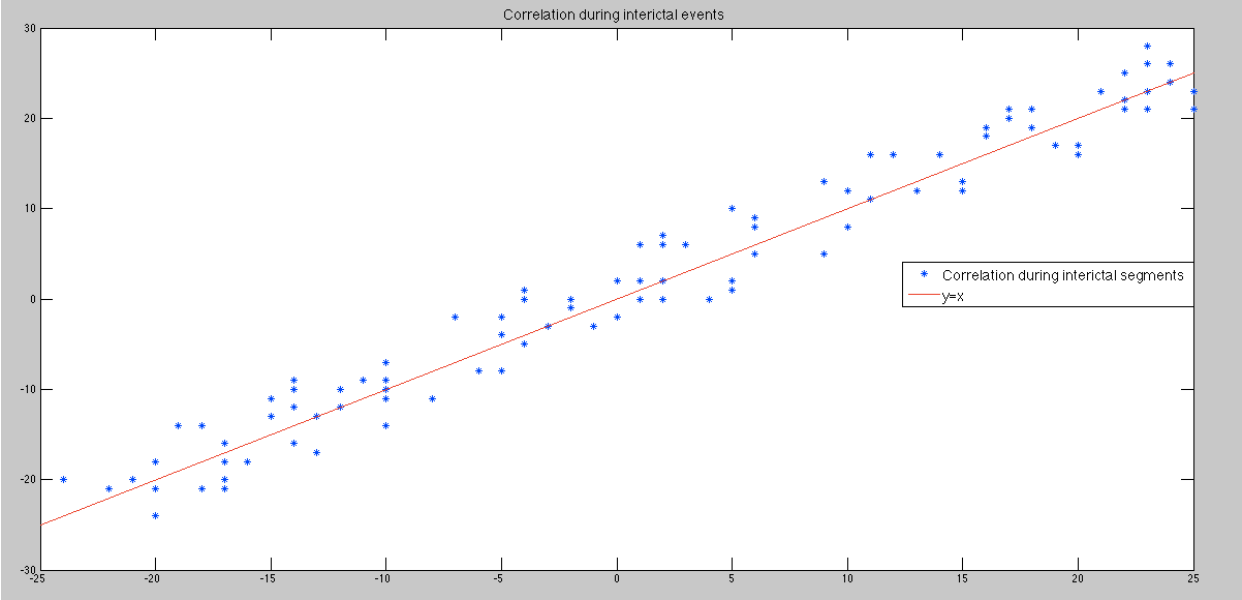

Figure 13: Correlation during interictal events

By contrast, when seizure occurs the correlation is like the one in Figure 14. The matrix in this case is:

$M = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ therefore eigenvalues are $\lambda = 1 \; and \; \lambda = 1$.

This means that during epilepsy attacks the energy is maintained almost constant in the representation of the eigenvalues spectrum.

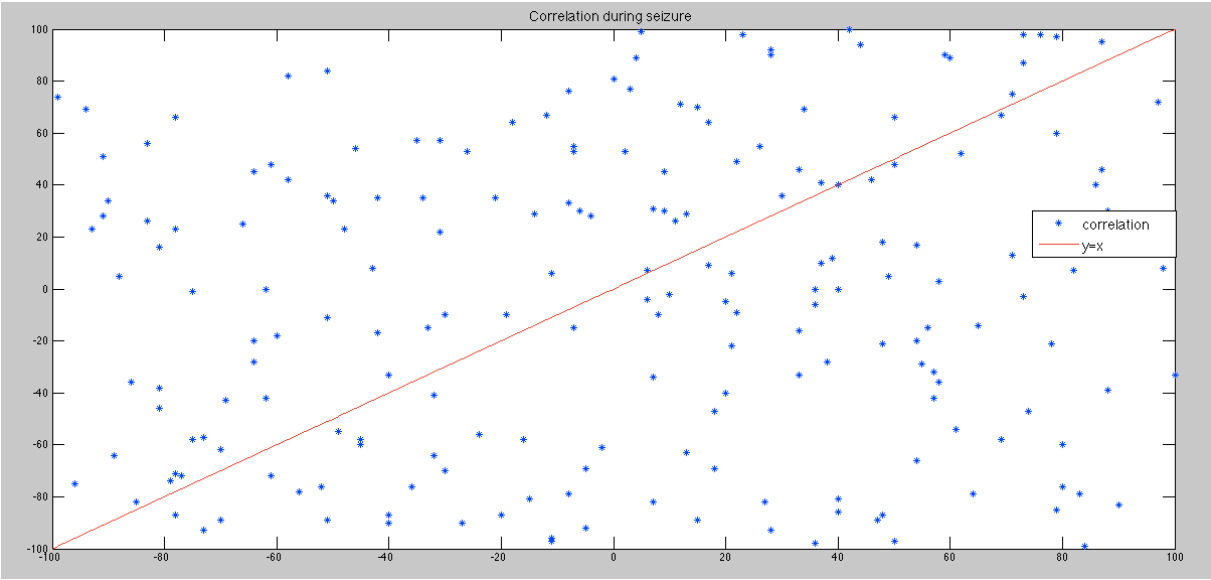

Figure 14: Correlation during epileptic seizures

Once we had the vector that concatenates the vectors of the variables (correlation, eigenvalues, PSD and power), a matrix that contains in each line the vector of variables of each of the segments of the training data (ictal and interictal) of one of the individual was created. This vector is what was called "**x**" during the explanation of the logistic regression (page 21). On the other hand, a vector that contain 1 or -1 depending if the segment is ictal (y=1) or interictal (y=-1) was created at the same time as "**x**". This vector is what it was called "**y**" during the explanation of the logistic regression (page 21).

These data is the one used to train the classifier with the logistic regression. In Matlab, there is one function that makes the logistic regression given these two vectors. This function is *glmfit* that is a generalized linear model regression in which you can determine the type of regression desired (in our case, *logit*, that is, logistic). Then, this function gives as output a value that is the vector "**w**". Once we had this vector, we take another matrix "**x**" that is the one that includes in each line the vector of variables of each of the segment of the test data. With that matrix "**X**" and the vector "**w**", we can calculate "**y**" and in consequence the categories of each segment of the test data.

Although this method is simple to use, it has one problem: *glmfit* calculates the vector "**w**" given the vector "**y**" and the matrix "**X**", the matrix "**X**" will be singular and the solution relies in including a new parameter that smooth the function as explained in Logistic Regression Section (page 21).

Thus, it was necessary to include this parameter λ in the piece of code. In the paper of Thomas Minka [12] all this process is explained in detail. In addition, he added a toolbox for Matlab with a wide variety of functions for logistic regression in which the function of the Newton's Method, that was used in order to extract optimized results of the algorithm, was included.

The process followed was the same as explained before with the Matlab function *glmfit* but in this case, the classifier was trained with this function. The function called *train_newton* takes as input the matrix "**X**", a starting guess parameter for "**w**" and the value of λ. Now, the matrix "**X**" is the multiplication of $\boldsymbol{x}^T \cdot y$. On the other hand, the initial value of **w** used is a vector of zeros and the value of λ will start on 1 and increasing it until the optimized algorithm is found.

4.1.2 PATIENTS

In the case of the patients, the process followed is the same. The only difference of this process is that, since the sampling frequency in some cases is much higher than in the case of the dogs, we need to decimate the signal to calculate the power spectral density in order to reduce the huge data and make a faster piece of code. Thus, decimation is a process by which the sampling frequency of a signal is reduced by a specific factor.

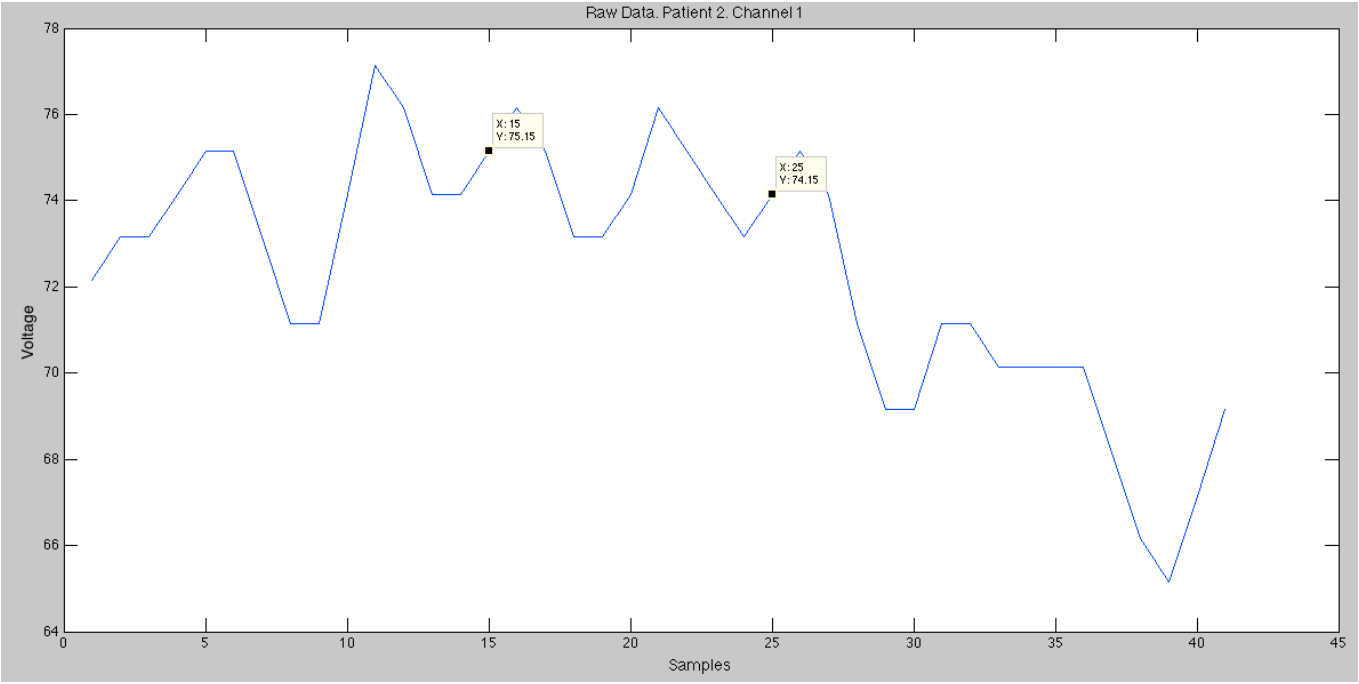

Figure 15: Raw Data. Patient 1. Channel 1

In our case, I have decimated the raw data by 10, that is, take one sample per 10 samples. The problem of that could be high changes in the amplitude of the signal that will be represented as overlapping in the Fourier transform and has as consequence the distortion of the power spectral density and therefore the distortion of the classifier algorithm. However, Figure 15 represents the raw data from sample 3520 to 3560 of the iEEG of patient 2 in channel 1. Taking into account that the maximum value of the whole raw data is 164.1 and the minimum is -201.9, we could say that the changes seen in Figure 15 between 10 samples are smooth so, decimation and in consequence overlapping will not be any problem in the calculation of the PSD. Moreover, in Figure 16 the PSD was represented and it is possible to see that all the values are accumulated at low frequencies and therefore, it also demonstrate that overlapping will not occur even with decimation 10.

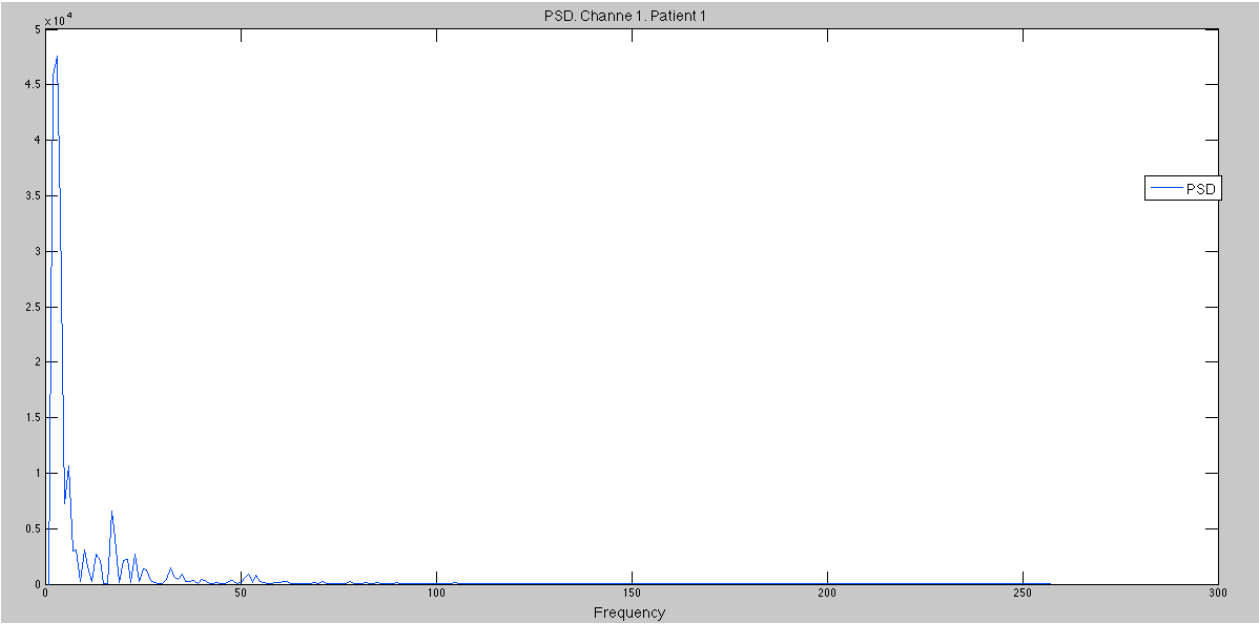

Figure 16: PSD of Patient 1

The rest piece of code is exactly the same as the one used for the dogs.

4.2 Early vs Late

The work in the second part was almost the same. The variables used are the same for the same reason and also in this section the data of the patients that have a higher sampling frequency is decimated. In addition, both *glmfit* and the toolbox of Thomas Minka were used to demonstrate that the problem continues to be a current issue in this part.

In this case, we need to differentiate interictal segments and late ictal segments (y=-1) from early ictal segments (y=1). As I have said before, early ictal segments are the ones that occur in the first fifteen seconds of the ictal segment while late ictal segments are the rest. This process could be done in two ways.

The first one consists in training the classifier with the ictal (both early and late) and interictal values and then applying the algorithm to the test values. However, this method has not very good results since it is very difficult to train the classifier to differentiate between late ictal and interictal segments in the

same category because they are very different between each other. So, it will have problems while differentiating them in the test files.

Thus, in order to solve that, a second method was developed. This process consists in training the classifier only with the ictal values (both early and late). In this way, then the algorithm is applied to the test values. This classifies all the segments into early (y=1) and late (y=-1) even when the segments are interictal. So, the next step is to multiply the result of this with the result of classifying the test segments into ictal and interictal. Thus, what we are achieving by doing that is that only the segments that are classify as ictal and early are classify as y=1 and the rest are classify as both late ictal or interictal.

In Figure 17 we could see the effect of the second method explained. The first graph represents the classification of the test data in ictal (y=1) and interictal (y=-1). Then, in the second plot the representation of the classification of early (y=1) and late (y=-1) when the classifier is only trained with ictal data. Finally, the third representation is the final separation between early (y=1) and late ictal/interictal segments (y=-1) of the test data that results from the multiplication of the values of the first graph and the results from the second one. Hence, the result of the classification with this method is much more efficient and realistic than the first method explained.

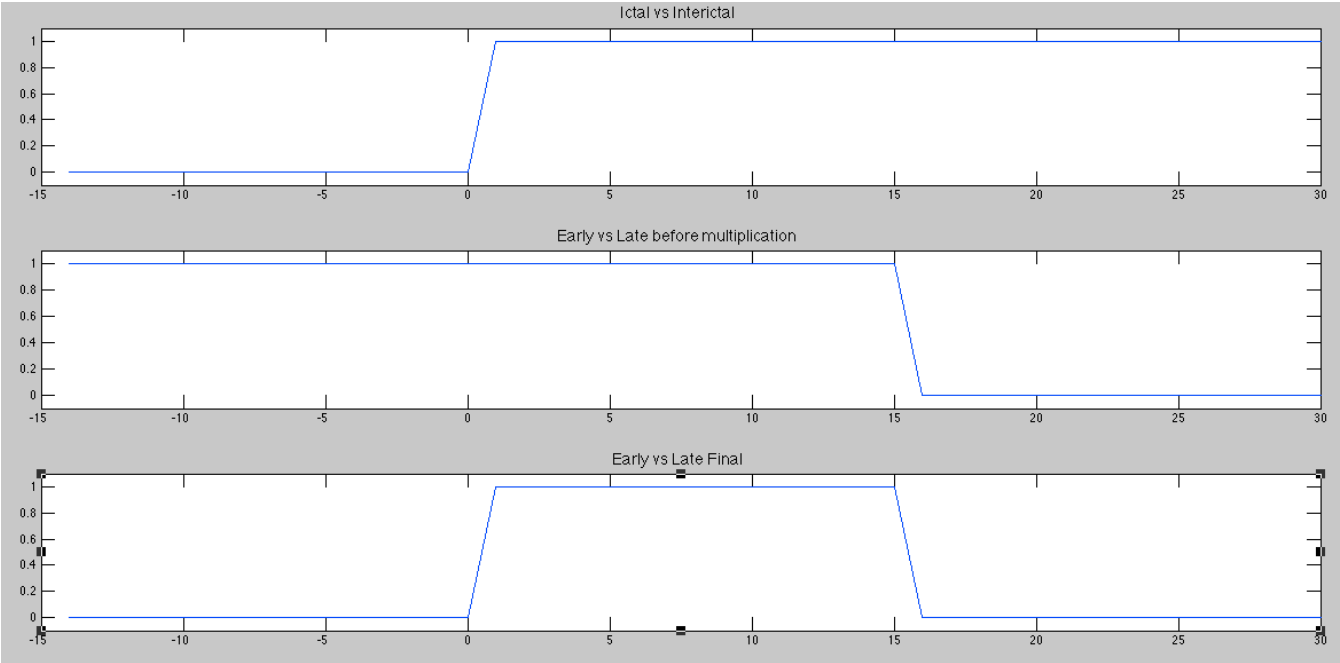

Figure 17: Effect of the algorithm

The main problem of this second method is that it is necessary to classify firstly the test data as ictal or interictal to allow me to then classify the same data as early or late. A simple scheme of the general ideas explained in the last paragraph is represented in Figure 18. In consequence to that relation between both support vector machines training, the algorithm is much slower and depends greatly on how good the classification of the data of ictal and interictal is.

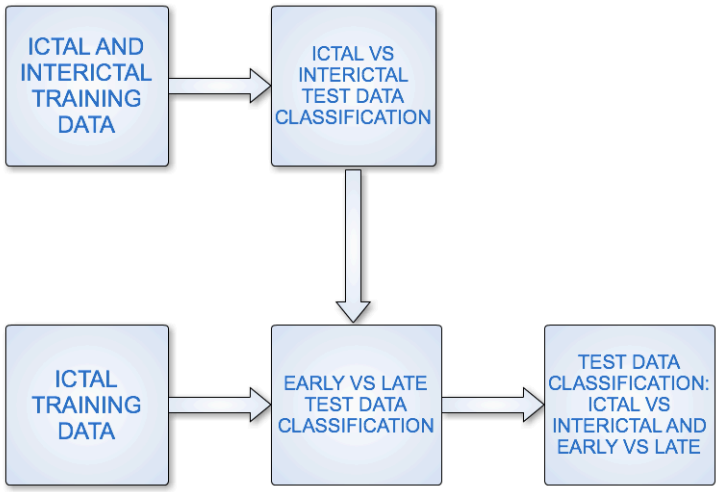

Figure 18: Diagram of general idea of classification

5. RESULTS

Although the development of the algorithm is straightforward, it is necessary to take into account how good the algorithm is. It is essential to make different estimations for different combinations of variables and different values of the parameter λ whose effect is explained in Figure 8. There exist a method very used in prediction and artificial intelligence that allow us to certificate in a proper way the validation of this machine learning that is cross-validation.

This process is used to make the validation of the algorithm applied to the classifier (for both classification: ictal vs interictal and early vs late). First of all, all the training data was randomly organized (both interictal and ictal). The main problem of this is that I am probably including some of the segments of each seizure attack in each of the fold. This will provide an overestimation of the results since we are classifying some ictal segments by training another segments that are quite similar to them. So, it is important to do a realistic and reliable validation to separate the ictal segments of each attack and include all of them in one single fold.

In this way, the cross-validation folding will depend on the number of epileptic attacks that each individual has in their training data (it is not constant in each individual. It varies from 2 to 12). Therefore, the final cross validation was done in the following way:

- First of all, the interictal training data was randomly organized and divided in the same number of subgroups as epileptic seizures has the individual.
- The ictal segments were divided into subgroups each including all the segments of each epileptic seizure. Then, each of these subgroups was added to one of the subdivisions of the random interictal subunits.

- Finally, what I have achieved is to have k folds where k represents the number of epileptic seizures and each of them including a random group of interictal segments and all the segments of each epileptic seizure.

By using this method, the overestimation explained before was avoided.

On the other hand, ROC complements the function of the cross-validation process for determining the quality of the results in each of the cases.

Both cross-validation and ROC space were followed with the combination of different variables in order to know which independent variables will provide the best algorithm that divides the files into ictal or interictal and early or late. Figure 19 exhibits the result in form of the ROC curve after making different combination. By training with all the independent variables (correlation, power, PSD and eigenvalues), the results is the better one followed by training without correlation; training without power; training without eigenvalues and finally the worst is the one that is trained without PSD.

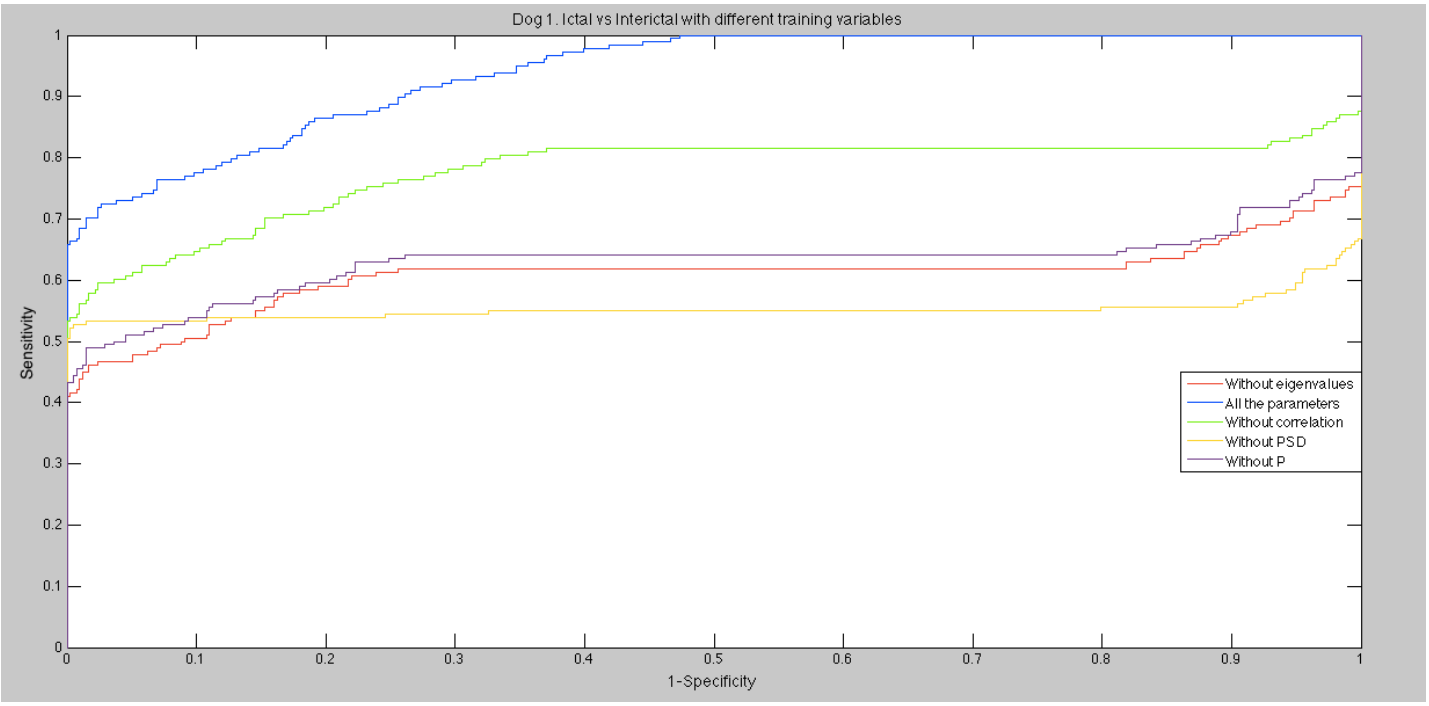

Figure 19: Comparison between training variables

5.1 Ictal vs Interictal

After applying cross-validation to the training data of the dog for the ictal vs interictal classification, the results are the ones shown in Figure 20. These results show the ROC curve of the prediction of ictal segments versus interictal segments in dogs.

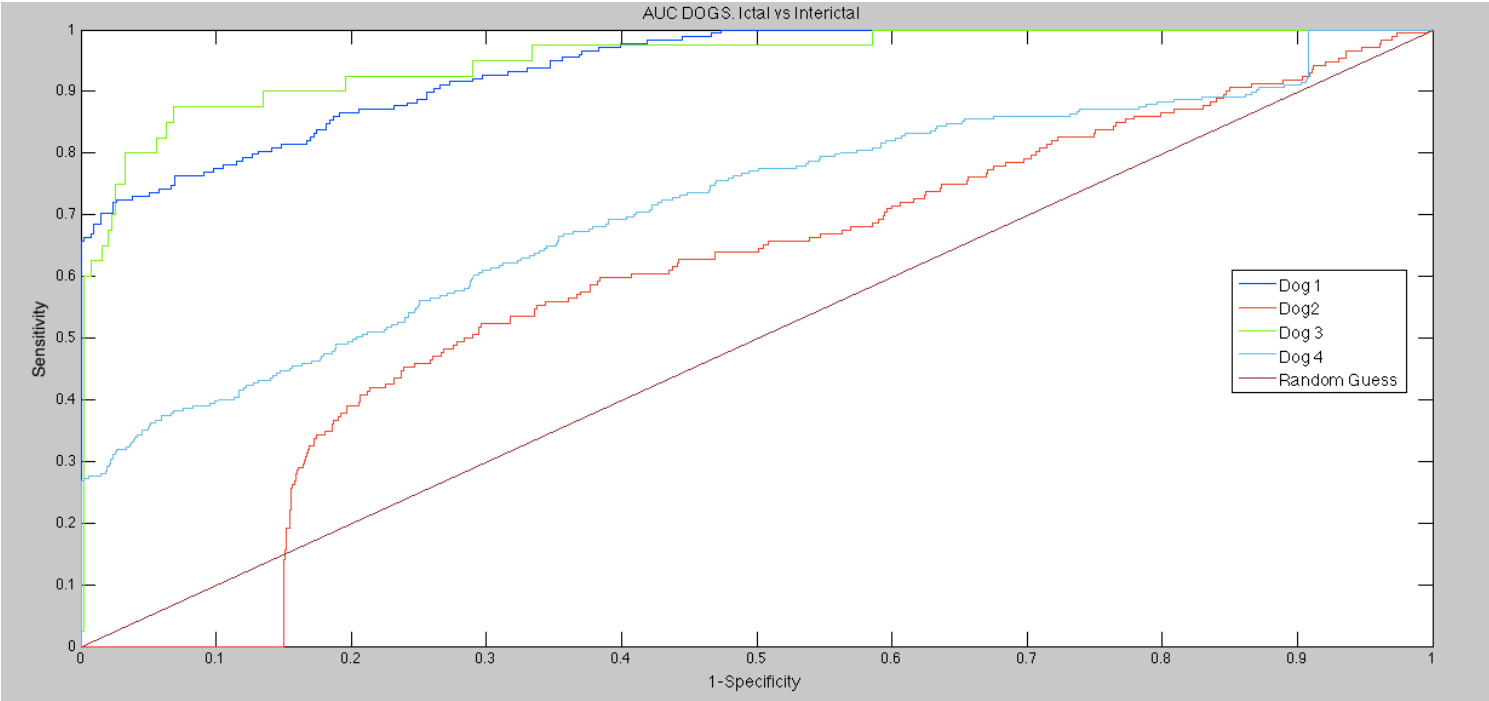

Figure 20: Ictal vs Interictal Dog Classification

In Figure 20, purple line shows the random guess and as I previously explained, the results must be above that line. This is not the case of Dog 2 which results are not the once expected. In order to search the possible error of that disconformity of the errors, I have make a ROC curve from each fold of the cross-validation that is shown in Figure 21.

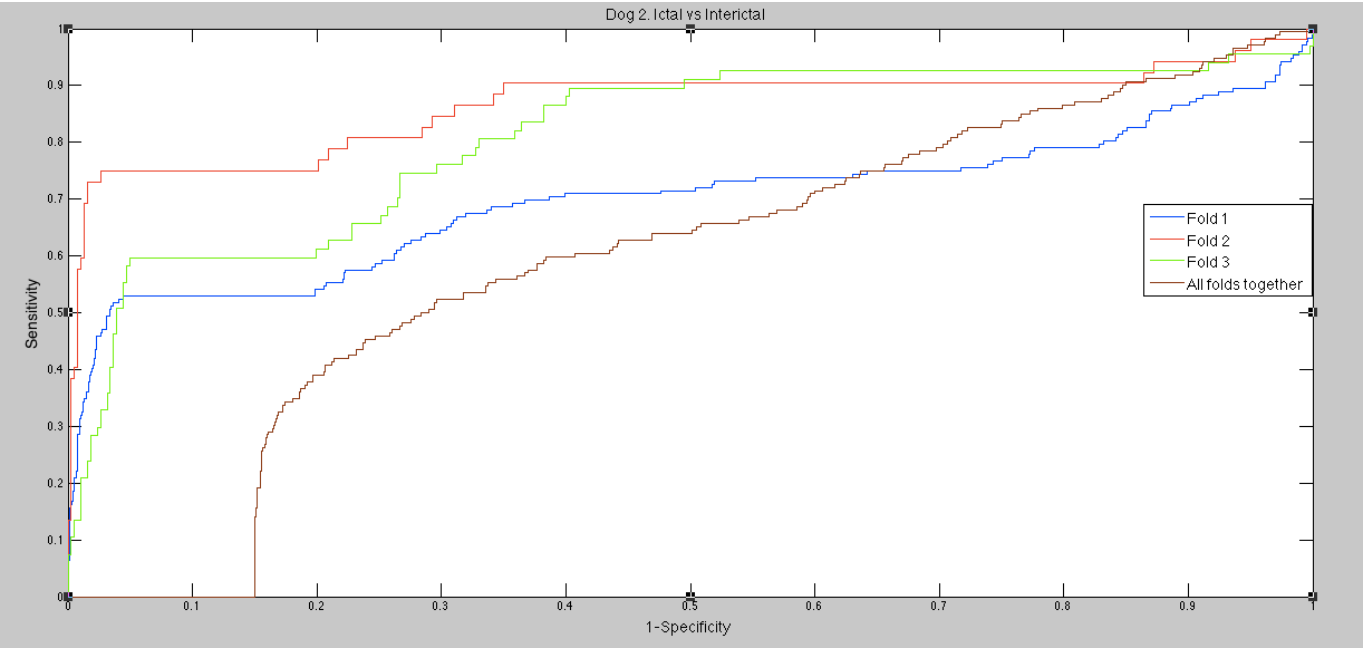

Figure 21. Comparison between ROC curve of 1 single fold of cross-validation and all the fold together

What these results show is that the results from each of the folds are better than when joining all of then making the ROC curve. The reason of that is that the threshold that limit the separation between ictal and interictal samples in each fold is different and when putting them together, they overlap and some values that previously were correctly separated, now are producing a worse result than by doing it separately.

For patients, the procedure followed is exactly the same as in the case of dog. The results of the patients are shown in Figure 22.

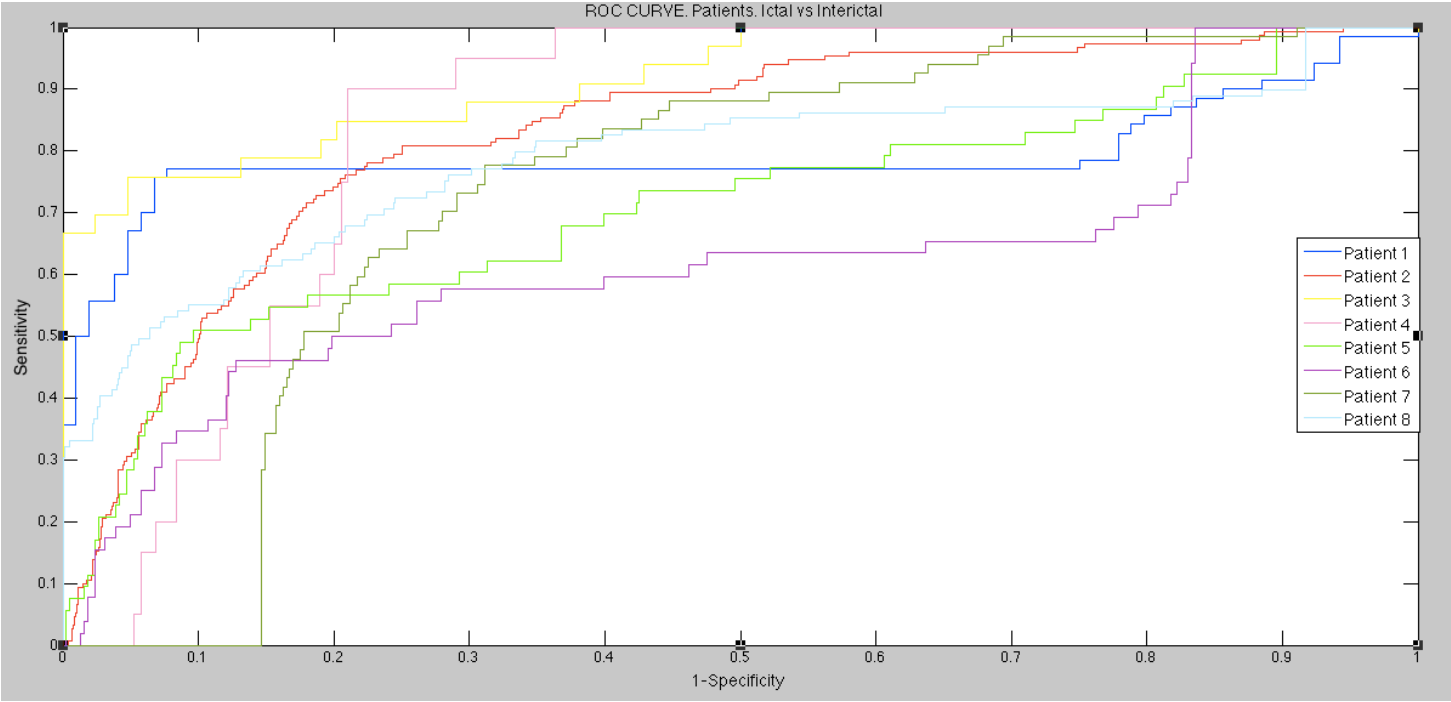

Figure 22: Ictal vs Interictal Patient Classification

The results of patient 7 and patient 6 are the worst between the eight patients and the explanation is exactly the same as the one given to the results of the dogs, that is, the different thresholds for each of the folds of the cross-validation.

SUBJECT	AUC Ictal vs Interictal
DOG 1	0.9374
DOG 2	0.6125
DOG 3	0.9508
DOG 4	0.7497
PATIENT 1	0.7896
PATIENT 2	0.8282
PATIENT 3	0.9127

PATIENT 4	0.8660
PATIENT 5	0.7103
PATIENT 6	0.6213
PATIENT 7	0.7345
PATIENT 8	0.7871

5.2 Early vs Late

The second part relies on the classification of early versus late. The application of these results is much higher because it will allow s to anticipate any event and provide a small time to act in response to that prediction. These results will be a bit worse than the ones before because they accumulate the errors of the classification of ictal versus interictal with their own errors of classification. This occurs because, since it was explained in Figure 17, the classification of early versus late depends on the classification on ictal vs interictal. The results show an AUC different for each individual.

SUBJECT	AUC Early vs Late
DOG 1	0.7564
DOG 2	0.5678
DOG 3	0.6187
DOG 4	0.7098
PATIENT 1	0.6193
PATIENT 2	0.6004
PATIENT 3	0.6178

PATIENT 4	0.7218
PATIENT 5	0.5635
PATIENT 6	0.5465
PATIENT 7	0.5619
PATIENT 8	0.6588

Figure 23 represents the AUC for different dogs for the problem of classification early vs late.

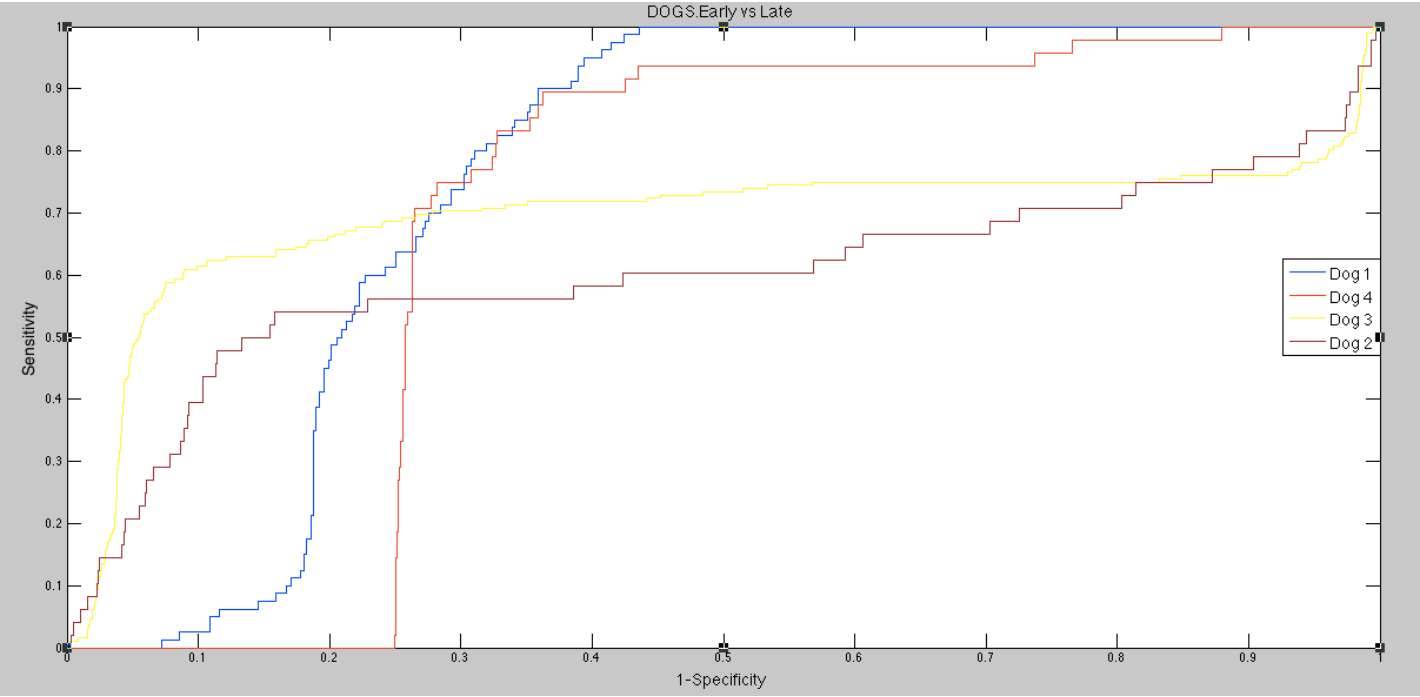

Figure 23: Early vs Late Dog Classification

Finally, the results of the classification of early versus late for patients are shown in Figure 24.

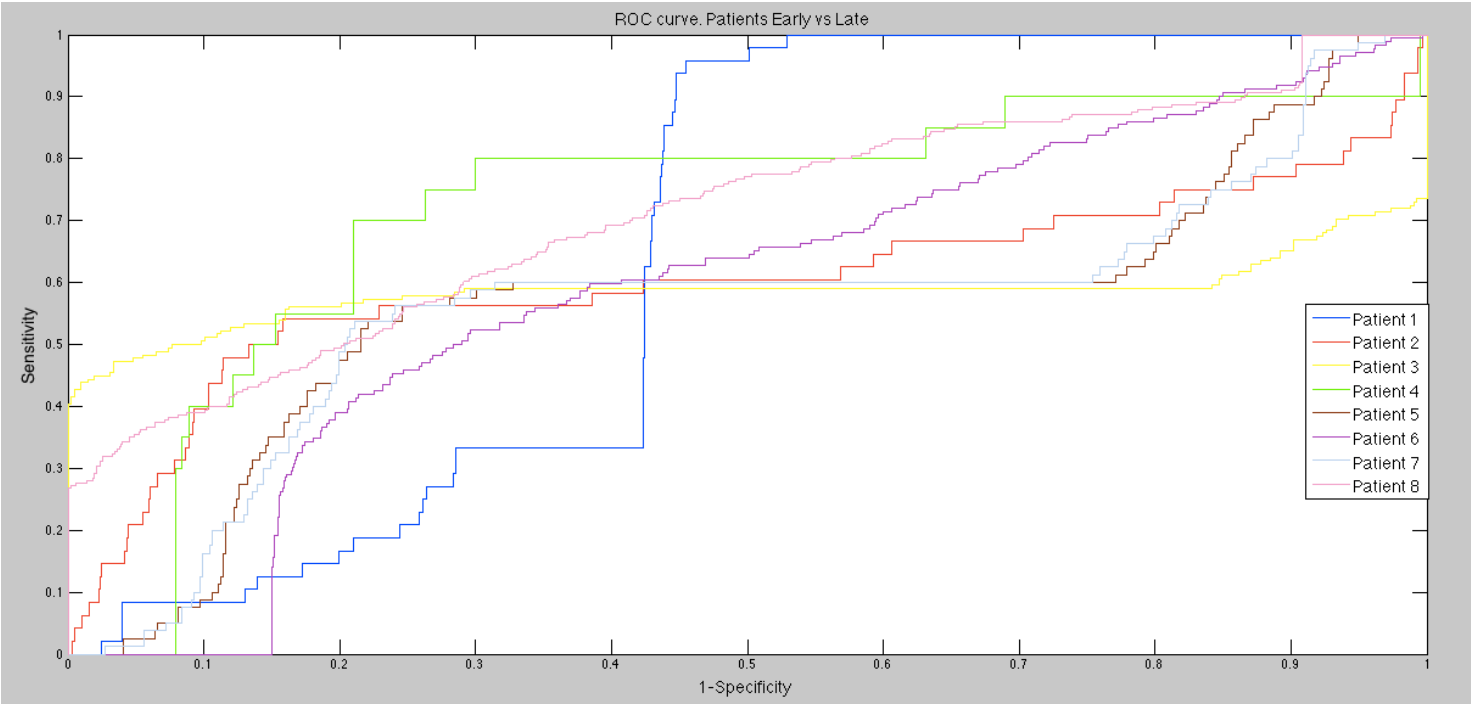

Figure 24: Early vs Late Patient Classification

As it can be seen in Figure 24, the results of the classification early versus late is worse in comparison with the ictal versus interictal. The main reason of that it the accumulation of errors due to the concatenation of both results but also because it is more difficult to separate the values of early and late since both are ictal segments and are quite similar to each other.

6. DISCUSSION

This project opens the possibility of forecasting of epileptic seizures from intracranial electroencephalogram by recording, annotating and analyzing large sets of continuous-recording data. It could be a definitive, final solution for those patients that do not show improvements or control with their drug delivery and also to reduce the side effects that these drugs present.

Although the first study of intracranial EEG in patients was done in 2013, comparing our results with the previous ones study with a Poison-process algorithm, they are much better. [4]

Some limitations of the previous studies were small data sets that imply a smaller number of seizures and interictal segments which make it more difficult to detect with a high efficiency the ictal segments. In our experiment, the data I am studying are continuous-recording data for very long period of times in comparison with the previous studies. In addition, the results are more reliable due to the method of evaluation (cross-validation) and the previous study of the independent variables.

However, the algorithm could be improved in future. Firstly, our classification method is one of the most simple and primary, so by a using a complex one the estimations will be better. It is essential to reduce the false positives that are the ones that make the classification to be worse and less efficient. Firstly, it is an important measurement to test the optimal bandwidth of both ictal and interictal segments. The bandwidth of the dogs in this project is 400 Hz but some recent investigations have shown that the EEG activity expands in a wider range (1000 Hz) [18]. Moreover, the algorithm only shows if an epilepsy attack is occurring but it does not display the channel or electrode where the ictal segment has a much higher activity.

In addition, the results could be very different for different patients. Those patients that show more or less periodical seizures are much easier to predict with this algorithm. By contrast, patients that show very different types of seizures or seizures that are developed faster are more difficult to predict and detect. In addition, it is much more difficult to classify correctly early versus late in comparison with ictal versus interictal that shows much better results. Thus, it is important to reduce false positives to have a better classification.

One more added problem to the algorithm is the slowness that it shows. The main reason of that is that the classifier is a non-linear kernel that in addition to the time, it consumes too much energy. Thus, one possible improvement of the future could be the possibility of make the kernel of the classifier linear but maintaining efficient results.

Although the independent variables or features used to train the algorithm are previously studied and tested, by adding some new features to the training algorithm, the results could also be improved.

Summarizing, this bachelor thesis opens the possibility of a future individualized and specialized treatment that provides a real-time continuous measurement and treatment, improving the patients' quality of life.

7. COSTS AND WORKING TIME

This final chapter will present the costs of the software needed for the project as well as the working time in the Bachelor Thesis and the corresponding estimation of the costs of having engineers working on that.

The main and essential components needed are a computer and Matlab software. The price of a computer is more or less 1000€ and the prize of the license of Matlab for a year is about 130€. However, the duration of the work was about 6-7 months, so the final costs are the following:

PRODUCT	PRIZE
Computer cost for 6 months	125€
Matlab license for 6 months	65€
TOTAL	190€

On the other hand, the prize of having engineers working on that would be approximately:

HOURS	COST PER HOUR	TOTAL COST
360	24 €	8640

Thus, the final cost is **8830 €.**

8. REFERENCES

[1] Peggy J Copple. The treatment of epilepsy: Principles and practices. *Archives of Pediatrics & Adolescent Medicine*, 148(7):769, 1994

[2]] Patrick Kwan and Martin J Brodie. Early identification of refractory epilepsy. *New England Journal of Medicine*, 342(5):314–319, 2000.

[3]P David Adelson, Edwin Nemoto, Mark Scheuer, Michael Painter, John Morgan, and Howard Yonas. Noninvasive continuous monitoring of cerebral oxygenation periictally using near-infrared spectroscopy: a preliminary report. *Epilepsia*, 40(11):1484–1489, 1999.

[4] J Jeffry Howbert, Edward E Patterson, S Matt Stead, Ben Brinkmann, Vincent Vasoli, Daniel Crepeau, Charles H Vite, Beverly Sturges, Vanessa Ruedebusch, Jaideep Mavoori, et al. Forecasting seizures in dogs with naturally occurring epilepsy. *PloS one*, 9(1):e81920, 2014.

[5] Kim Ann Zimmermann. Nervous system: Facts, function and diseases, 2015. URL www.livescience.com/22665-nervous-system.html

[6] Gerard J Tortora and Bryan H Derrickson. *Principles of anatomy and physiology*. Wiley, 2011.

[7] John E Hall. *Guyton and Hall textbook of medical physiology*. Elsevier Health Sciences, 2010.

[8] World Health Organization. Epilepsy, 2013. URL http://www.who.int/mediacentre/factsheets/fs999/en/

[9] Charles R Newton and Hector H Garcia. Epilepsy in poor regions of the world. *The Lancet*, 380(9848):1193–1201, 2012.

[10] Carl French. *Data Processing and Information Technology*. Thomson, 10 edition, 1996.

[11] Bishop, Christopher M. *Pattern recognition and machine learning*. Springer, 2006.

[12] Thomas P Minka. Algorithms for maximum-likelihood logistic regression. 2003.

[13] Refaelizadeh, Payam; TANG, Lei; LIU, Huan. Cross-validation. *Encyclopedia of database systems*. Springer US, 2009. p. 532-538.

[14] Flach, Peter A. The geometry of ROC space: understanding machine learning metrics through ROC isometrics. *ICML*. 2003. p. 194-201.
[15] Uppen and Mayo Clinic. Upenn and mayo clinic's seizure detection challenge, 2014. URL kaggle.com/c/seizure-detection.
[16] Yun Park, Lan Luo, Keshab K Parhi, and Theoden Netoff. Seizure prediction with spectral power of eeg using cost-sensitive support vector machines. *Epilepsia*, 52(10):1761–1770, 2011.
[17] Maurice Stevenson Bartlett. Smoothing periodograms from time series with continuous spectra. *Nature*, 161 (4096):686–687, 1948.
[18] GA Worrell, K Jerbi, K Kobayashi, JM Lina, R Zelmann, and M Le Van Quyen. Recording and analysis techniques for high-frequency oscillations. *Progress in neurobiology*, 98(3):265–278, 2012.